

SYSTEM AND METHOD FOR PERFORMING ANALYSIS ON WORD VARIANTS

ABSTRACT OF THE DISCLOSURE

5 A computer-readable medium stores a first lexicon data structure for
lexicon words. The first data structure includes a host form variant field
containing a host form variant such as a clitic host form variant, a host form
field containing the host form of the host form variant (only present if the forms
differ) such as a clitic host verbal form, and a verification field indicative of
10 whether the host form variant is a valid word. The first data structure also
includes a segment association field containing data or segmentation bits
associating the host form variant with certain types of attachment entries in the
lexicon, which also contain data or segmentation bits, to define valid
combinations between the host form variant and at least one of the attachment
15 entries in the lexicon. A second lexicon data structure for each of the attachment
entries in the lexicon is also stored.